Our task is to predict which customers will churn in next 6 months based on the given information of customers

Approach:

Our target variable is highly imbalanced. So we included synthetic samples to balance the data using smote

Then we used LGBM classifier through which we got to know the important features as Age, Balance and Income

So we added new features based on Age, Balance and income through which we were able to improve our score a little

Then we combined Product holdings which is the number of products a customer is having and credit card into single feature. Then we dropped those two. Through this also we were able to improve our score a little

We tried combining many features and taking mean of features but it didn’t help in score improvement so we dropped those

We also tried XGB, Catboost algorithms as well which gave decent score. But we got best score on leaderboard with LGBM so sticked on to that and tried tuning hyperparameters. We also checked by combining LGB and XGB . Though score improved in notebook, there is no improvement in leaderboard score.

Our Final model is based on LGBM with tuned parameters and 10 fold cross validation with which we were able to get score of 0.6003 in leaderboard